SELF-ASSEMBLING GAMES AND THE EVOLUTION OF SALIENCE

JEFFREY A. BARRETT

Abstract. We consider here how a generalized signaling game may self-assemble as the saliences of the agents evolve by reinforcement on those sources of information that in fact lead to successful action. On the present account, *generalized signaling games* self-assemble even as the agents coevolve meaningful representations and successful dispositions for using those representations. We will see how reinforcement on successful information sources also provides a mechanism whereby simpler games might compose to form more complex games. Along the way, we consider how an old game might be appropriated to a new context by reinforcement on successful associations between old and new saliences.

1. Self-Assembly Through the Evolution of Salience

Our assessment of the world and how it is constructed depends on what we notice. Inasmuch as it is perceptible, a feature of the world may potentially affect our deliberations and actions. Of course, there are a great many things that are in principle perceptible but not in fact noticed. Such things do not affect our deliberations or actions—they are not salient.

What we know and how we deliberate depends on what is salient. As a consequence, the evolution of salience is a central issue in both epistemology and rational choice. Regarding the latter, Thomas Schelling ([1960]) used the notion of focal points to address the problem of equilibrium selection in strategic deliberation. Along similar lines, David Lewis ([1969]) appealed to preexisting saliences to explain how conventions might be established in the context of classical game theory. One of the virtues of evolutionary game theory is that it allows one to explain how saliences might *evolve naturally* on repeated plays of a game where each play may affect one's behavior on future plays. Skyrms ([2010]) showed that one need not assume preexisting saliences to evolve signaling conventions in the context of an evolutionary version of the Lewis signaling game. Rather, effective saliences that break symmetries between alternative strategies may evolve under a simple dynamics like reinforcement learning or the replicator dynamics. Here saliences evolve to tie strategies to states of nature. Barrett ([2007)]) and Purves ([2018)] have shown how such evolved saliences may affect an agents metaphysical commitments by determining how she individuates objects and properties in the world.

We are concerned here with how saliences may evolve to structure our interactions and reflections more generally. They accomplish this by assembling games. Specifically, we will consider how *generalized signaling games* may self-assemble by means of evolved saliences. The aim is to shed light on how our perceptual and cognitive systems may evolve the capacity to focus on those aspects of the world that matter, use information gained from these for successful action, and evolve increasingly complex forms of inference, representation, and communication by means of the evolution of further saliences.

Lewis-Skyrms signaling games and their variants have been well-studied, but less has been done to explain how such games might come to be played in the first place and how the games themselves might evolve over time.[1] Here we use the theory of self-assembling games to explain how generalized signaling games might evolve from more basic interactions between agents.[2]

Here we investigate how a generalized signaling game might self-assemble by simple reinforcement. In brief, reinforcement on past success may forge a network of saliences that structure the evolving generalized game.[3] The key idea is that a generalized signaling game might self-assemble as agent saliences evolve by simple reinforcement to track those sources of information that in fact lead to successful action. And, as the game self-assembles, the agents often coevolve the ability to play it optimally.

Generalized signaling games may take many forms. When successful, such games might be understood as evolving rules or, more generally, as concrete implementations of recursive functions. Explaining how such games self-assemble, then, explains how rule-following and general algorithms might evolve.[4] Depending on how one understands cognition, explaining how an algorithm might evolve by self-assembly may go some way in explaining a how a cognitive system might self-assemble by simple reinforcement to perform epistemic functions like representation, communication, and inference.

---

[1]Lewis-Skyrms signaling games are described in (Lewis [1969]) and (Skyrms [2010]). The former involves classical games, the latter evolutionary games. For examples of variants and applications see (Barrett [2007], [2014], [2016]; Skyrms [2008], [2010]; Barrett *et al.*[2018], [2019]). For a broad spectrum of alternative dynamics for such games see (Barrett [2006]; Barrett and Zollman [2009]; Alexander *et al.* [2012]; Huttegger *et al.* [2014]; Barrett *et al.* [2017]). And see (Argiento *et al.* [2009]) for an analysis of the limiting properties of simple reinforcement learning in the simplest variety of signaling game.

[2]See (Barrett and Skyrms [2017]) for an introduction to self-assembling games.

[3]This is the same mechanism used to explain how epistemic networks might self-assemble in (Barrett *et al.*[2019]). Indeed, the present paper might be thought of as applying the key ideas from (Barrett *et al.* [2019]) to the theory of self-assembling games from (Barrett and Skyrms [2017]). See (Pemantle and Skyrms [2004]) for another model of network formation by reinforcement.

[4]See (Barrett [2014]) for a discussion of how rule-following might evolve from agents playing an evolutionary game. Here we are concerned with both how the game itself and how the rule it comes to implement might coevolve.

Here we consider three simple evolutionary models. In the *signaling model*, a generalized signaling game self-assembles as the agents coevolve systematically interrelated saliences and a meaningful language. The players learn what sources of information matter for successful action even as they learn how to react to the specific content of those sources. In the *template transfer model*, agents learn to appropriate an old evolutionary game to a new context by reinforcement on successful associations between old and new saliences. In doing so, they evolve an analogy that allows them use old evolved dispositions to accomplish a new task. In the *modular composition model*, the agents learn to compose simpler games into a more complex game by reinforcement on successful sources of information. The evolution of new saliences may allow the composite system to evolve successful dispositions for the complex game by composition more efficiently than it might have evolved them from scratch.

Each of the models involves only primitive resources. Simple reinforcement has a long history of use in modeling both animal and human learning and is an extremely low-rationality learning dynamics.[5] The agents here begin with only simple reinforcement, then bootstrap to more sophisticated representational and computational abilities.

Importantly, the agents in these models might be understood as separate individuals or as functional units of a single individual. As a result, stories along the lines of those told here may be told in other evolutionary contexts and at other levels of description.[6]

2. A Self-Assembling Signaling Game

The first model illustrates how a generalized signaling game might self-assemble as the agents' saliences coevolve with their dispositions to signal and act. As with the other models we will discuss, the signaling games here self-assemble in the context of common-interest between the interacting agents.

Consider two senders a and b who have access to four states of nature (randomly selected on each play from 0, 1, 2, or 3 in an unbiased way) and one receiver r who has access to the behavior of both senders and who performs one of four actions (either 0, 1, 2, or 3). On each play of the game, both senders see the full current state of nature, then each sends one of two possible signals (either 0 or 1). The receiver may pay attention to the signal produced by sender a or may pay attention to the signal produced by sender b or may pay attention to the signals produced by both senders $a\&b$. The receiver then performs an action that is successful if and only if it matches the current state of nature (each action only matches the correspondingly-labeled state). If the receiver's action is successful, then all players are rewarded.[7]

[5]See (Herrnstein [1961]) for an early and influential example of the study of simple reinforcement learning.

[6]This point matters immediately. There are four states of nature that are naturally salient to each the two senders in the self-assembling signaling game we consider in the next section. Similarly, each sender has two actions that are salient to the receiver. Such initial saliences may be accidental or how these state types came to be salient to the senders and how these action types came to be salient to the receiver might be explained, in a particular concrete context, in much the same way that we explain here how the signals of *both* senders come to be salient to the receiver.

[7]See (Barrett [2007]) for an early version of this game where the receiver's saliences are simply stipulated. The present model shows how the simpler evolutionary game might itself evolve.

The senders' signals do not mean anything initially. And, since each sender has only two possible signals, neither sender alone can evolve signals that represent the four possible states of nature. At least initially, they do not even know which aspects of nature to attend to divide their representational labor. To be successful, the senders must coevolve dispositions to attend to different, systematically interrelated, aspects of nature even as they learn how to send signals that represent what they see when they do.[8] The aspects of nature that each sender learns to watch to determine her signals are that sender's evolved saliences. Similarly, which senders the receiver learns to watch are his evolved saliences. Putting the pieces together, to be successful, the senders' saliences and the signals they choose to send must coevolve with the receiver's saliences and the actions they choose to perform—and here that means that the senders' signals must evolve meanings that are systematically interrelated in such a way that they *together* represent the four states of nature *and* the receiver must coevolve the disposition to pay attention to the signals from *both senders* and use those signals to act in a way that corresponds to their evolved meanings.

We will suppose that the agents learn by simple reinforcement. On each play of the game, the state of nature is randomly determined with unbiased probabilities. Each sender then sees the current state and randomly draws a ball from her corresponding signal urn. We will suppose that each of the senders' signal urns starts with one ball of each possible signal type 0 or 1. The ball drawn by each sender determines that sender's signal.

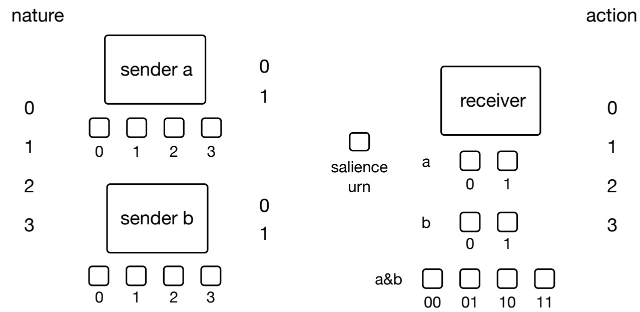
The receiver starts by drawing a ball from his salience urn. This tells him which sender to watch. We will suppose that the salience urn starts with one ball of each type a (watch sender a), b (watch sender b), and $a\&b$ (watch both sender a and sender b). If the receiver draws the ball to watch sender a, he draws an action ball at random from his sender a urn that matches the signal from a. If he draws the ball to watch sender b, he draws an action ball at random from his sender b urn that matches the signal from b. And if he draws the ball to watch both sender a and sender b, he draws an action ball at random from his sender $a\&b$ urn that matches the *pair of signals* produced by both senders. We will suppose that each of these action urns initially contains one ball of each action type $0, 1, 2$, and 3. The receiver then performs the action indicated by the ball drawn.

If the receiver's action matches the current state, it is successful and the senders and receiver return the balls they drew to the urns from which they were drawn and add a new ball to that urn of the type that led to the successful action. Otherwise, they simply return the balls they drew to the urns from which they were drawn.[9]

On simulation, the senders begin by signaling randomly, the receiver watches just sender a, just sender b, or both senders at random, and the receiver subsequently acts randomly and, hence, the agents succeed only by chance. On repeated plays, however, the senders often evolve *systematically interrelated saliences* that allow them to represent all four of the states of nature in their joint signals and the receiver coevolves the disposition to pay attention to the actions of *both* senders.

[8]If successful, the senders will evolve effective natural kind terms as they evolve saliences that determine what they attend to in producing their signals. See (Barrett [2007]; Purves [2018]) for discussions of the evolution of natural kinds.

[9]Note that all agents reinforce on success, not just those in fact involved in the successful action. The thought is that the senders may not know who is being watched on a particular play. That said, this does not seem to matter much for the general behavior of the game.



With 1000 runs and 10^6 plays per run, this occurs on 0.47 of the runs. A run counts as successful if the composite system is found to have a cumulative success rate better than 0.80.[10] Indeed, the cumulative success rate on such runs is typically better than 0.97. For the purpose of comparison, if the receiver starts with the *hard-wired disposition* to watch both senders, about 0.73 of the runs have cumulative success rates over 0.80 with 10^6 plays per run. So while the process is somewhat slower and less reliable when the receiver has to learn which senders to watch, it nevertheless succeeds in evolving an optimal signaling system about half of the time.

The senders evolve different representational roles and the receiver evolves to use information from both senders in one way or another on about 0.77 of the runs (this is the proportion of runs where the composite system exhibits a cumulative success rate better than 0.60, a rate that is impossible given the representational resources of just one sender). About half of the time, the receiver evolves to watch both senders and the agents jointly evolve an optimal signaling system. But an additional quarter of the time suboptimal things happen that also involve the receiver processing information from both senders. Among these, the receiver sometimes evolves to play a mixed strategy of watching both senders on some plays and watching just one select sender on other plays. Here the select sender's signals evolve to serve each part of the receiver's mixed strategy. Here the agents exhibit a cumulative success rate in the gap between 0.50 and 0.75. On other runs, the receiver focuses on a single sender and evolves a signaling system with just that select sender. In this case, since the select sender has only two signals, there is an information bottleneck and the cumulative success rate never does better than 0.50. Importantly, on 1000 runs with 10^6 plays per run the composite system is always found to have a cumulative success rate better than 0.45, so whatever happens, the agents nearly always evolve dispositions that do significantly better than chance.

At the finest level of description, this game involves four interacting agents: the two senders, the salience decider, and the receiver. The two senders each have sixteen pure strategies, there are three saliences, and the receiver has either four or sixteen depending on the chosen salience for a total of 6,144 pure strategy profiles. There are also the potentially infinite collection of associated mixed-strategy profiles. While a full equilibrium analysis would require significant bookkeeping, one

[10]The 0.80 cutoff works well in identifying optimally successful runs as the most successful sub-optimal pooling equilibrium has a success rate of 0.75. Longer runs do slightly better against this cutoff. The cumulative success rate of 0.52 against the 0.80 cutoff is observed on 10^7 plays per run.

can pin down a few points to compare with the simulations. Among the theoretically possible pure equilibria are profiles where the senders babble and the receivers action is independent of their signals for an expected success rate of 0.25, profiles where one of the senders signals divides the states into two types and the receiver only looks at that sender for an expected success rate of 0.50, profiles where the two senders signals fail to perfectly cross-cut the four states of nature yet each carries some information and the receiver watches both senders for an expected success rate of 0.75, and profiles where the two senders signaling perfectly cross-cut the four states of nature and the receiver watches both senders for perfect signaling.[11]

No examples of the 0.25 babbling equilibria were observed on simulation. One does see runs corresponding to one-sender 0.50 equilibria and two-sender 0.75 equilibria. One also sees runs with cumulative success rates in the gap between 0.50 and 0.75. Importantly, the plurality of runs corresponded to perfect signaling equilibria. The runs with cumulative success rates in the gap between 0.50 and 0.75 often involve the receiver mixing over watching one sender and watching two senders. Of course, to have a cumulative success rate better than 0.50 the receiver must at least sometimes watch *both* senders and must know how to succeed better than half of the time when he does. Instead of mixing on the signal source, one might imagine that the receiver would do better by always watching both senders.[12] But, if this happens on reinforcement learning, it happens very slowly.

On each run of the game a signaling game self assembles as the receiver evolves a network that determines the probability of his consulting each sender. Whether the game that ultimately evolves turns out to be a one-sender or two-sender game depends on the evolved network. Along the way, the senders and receiver coevolve signaling dispositions that exploit the evolving network. Put another way, the signaling game self-assembles by reinforcement on the receiver's saliences even as the agents coevolve optimal strategies for playing that very game. And, as we have seen, a network structure that allows for *optimal signaling* often self-assembles even as the agents together coevolve optimal signaling dispositions given the coevolving structure of the game.

Another version of this model is instructive. Consider the same model but suppose that one of the two senders, say sender B, has four signal types at her disposal, and that the receiver can condition on each of these if he chooses. Sender A, again, has just two signals at her disposal.

In this case one might imagine that the receiver would learn to watch just sender B since she can in principle communicate all of the states of nature with her four signals and the pair of agents might then evolve an optimal signaling system. Indeed, sometimes this happens, but there are many other things that happen, and the behavior of the agents can be subtle. Sometimes the receiver evolves to watch both senders or a mixture of one and both. Sometimes when the receiver watches both senders, the senders' signals together code for the current state of nature as in the two-sender model where each sender has only two signals. Other times, the receiver watches both senders, but some (or even all) signals represent states on

[11] See (Barrett [2007]) for a discussion of a simpler model that exhibits these two-sender 0.75 pooling equilibria.

[12] This is clearly the case if the salience decider randomly mixes between watching one sender and watching two in a way that does not depend on the conditions under which each sender is reliable.

their own. And in each case, there can be some evolved redundancy in the meanings individual signals and signal pairs.

On 1000 runs with 10^6 plays per run, this version of the model exhibits a cumulative success rate 0.84. This is better than a system where each sender has two signals and the receiver is hardwired to watch both (that has a cumulative success rate of about 0.73 with the model parameters here (Barrett [2006])). But, importantly, it is also better than a system where one sender has four signals and the receiver is *hardwired to watch just that sender* (that has a cumulative success rate of about 0.78). The senders here jointly have eight signal pairs when they only need four signals total to represent the states of nature. The high run success rate means that the receiver is evolving a network that exploits these extra signaling degrees of freedom to avoid suboptimal pooling equilibria. That is, the receiver is not just evolving saliences that track the information he needs for successful action—he is evolving saliences that compensate for inefficiencies in simple reinforcement learning.[13]

Reinforcement learning also allows for old evolved dispositions to be appropriated to a new context. In the next section we consider how agents might coevolve the ability to represent truth values and to use them to perform logical operations. In the two sections following that we will turn to consider how agents might evolve new saliences that allow such pre-evolved dispositional templates to be appropriated to new contexts or combined to provide stronger computational abilities more efficiently than those abilities might evolve from scratch.

3. The Evolution of Logical Operations

Agents may coevolve the ability to represent facts and to apply logical operations in the context of a generalized two-player signaling game under simple reinforcement. As an example we will consider how the logical operation *nand* may coevolve with signals that come to represent truth values in a two-player one-receiver signaling game.[14]

On each play of the game nature randomly determines the truth values of two propositions P and Q with unbiased probabilities. The two senders again each have two possible signals 0 and 1, but here one sender has access to the truth value of P and the other to the truth value of Q.[15] The receiver has access to the signal from each sender. On each play of the game, he performs an action T or F that is successful if an only if the action matches the truth value of P *nand* Q.

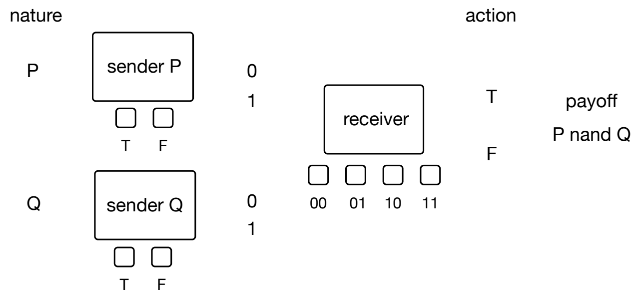
We will suppose that the agents learn by simple reinforcement and that each urn starts with one ball of each relevant type. The agents' actions are determined by the balls they draw from the urns corresponding to what they see. When successful, each agent adds a ball of the successful type to the urn from which she drew.

On simulation, the *nand* game typically *coevolves* a basic signaling language and the logical operator *nand*. On 1000 runs with 10^6 *plays/run*, 0.71 of the runs

[13]Having more terms available than what the agents need helps them avoid suboptimal pooling equilibria in Lewis signaling games (see (Barrett [2006]; Skyrms [2010]) for discussions). What makes the present case remarkable is that the receiver is coevolving saliences that take advantage of this effect.

[14]See Skyrms ([2008]) as a guide to early work on signaling games that also carry out logical inferences. See also (Barrett and Skyrms [2017]; LaCroix [2020]).

[15]Again, such initial saliences may be accidental or may have evolved from earlier interactions in the way that the evolution of specific saliences is explained here.



are found to have a cumulative success rate of better than 0.80, 0.62 of the runs better than 0.90, and 0.50 of the runs better than 0.95. Sometimes the game fails to evolve the operator *nand*. When it fails, the agents get stuck in a suboptimal partial-pooling equilibrium with a cumulative success rate of just under 0.75.

Agents playing a similar game may evolve the operator *or* from scratch by simple reinforcement just as *nand* evolves here. On simulation, this evolution is precisely as fast and effective. But if one has already evolved *nand*, *or* may evolve significantly faster by the process of template transfer.

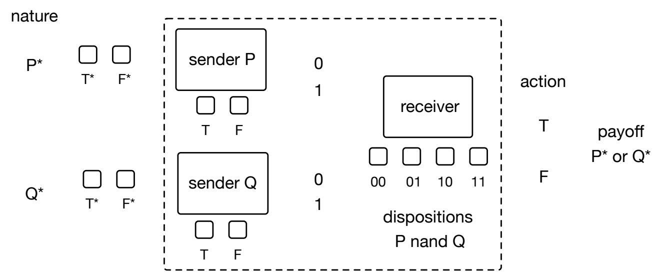
4. Template Transfer

Template transfer involves a set of old evolved dispositions being appropriated to a new context. This might occur by means of reinforcement on successful associations between old and new saliences. In the present model we suppose that two senders and one receiver have already evolved the dispositions to represent the truth values of P and Q and to compute the truth value of P *nand* Q. This system is represented by the dashed box in the figure below. One might think of this as a pre-evolved logical template. The template dispositions are determined by the agents' evolved saliences and the dispositional contents of their urns. For the purposes at hand, we will suppose that these dispositions are relatively fixed and focus on how the template might be appropriated to a new context.[16]

On each play of the template transfer game, nature randomly determines the truth value of two new propositions P^* and Q^*. Each old sender is given two translation urns that each start with one ball representing each of the corresponding sender's old signaling urns. The P-sender's T^* and F^* translation urns, for example, each start with one T ball corresponding to one old signaling urn and one F ball corresponding to the other. The P-sender sees the truth value of P^*, then draws a ball from her corresponding translation urn. The ball drawn from the *new* translation urn tells the sender which *old* signal urn to draw a ball from to determine her signal to the receiver. The Q-sender does the same thing to determine her signal given the input from Q^*. Here the receiver's action is successful if and only if it matches the truth value of P *or* Q. If the action is successful, then each sender adds another ball of the type drawn from the new translation urn to the urn from which she drew on this play of the game.

---

[16]The assumption is that the evolution of template transfer is much faster than the coevolution of the dispositions coded for in the template. To model this we will simply fix the template dispositions here.



On simulation, the template transfer game typically appropriates the old *nand* rule to a new context to produce the behavior of *or* an order of magnitude faster than *or* might evolve on the same reinforcement dynamics from scratch.[17] Specifically, on 1000 runs with 10^5 plays/run, 0.78 of the runs exhibit a cumulative success rate of better than 0.80, 0.61 of the runs better than 0.90, and 0.50 of the runs better than 0.95 (compare these to the results in the from-scratch game above). The agents evolve to map the new inputs to the old signals in a way that exploits the fact that the receiver's dispositions have already evolved to calculate *nand* on the old signals. This works because both *or* and *nand* produce F on exactly one pair of inputs. More generally, all sixteen binary logical operators might evolve quickly from any set of five operators that represent each number of F's that might be produced by a binary operator from zero to four.

The composite system evolving to map the new truth-functional inputs to the pre-evolved *nand* template in such a way as to mimic *or* in the novel context might be understood as a form of analogical reasoning. Specifically, the agents here discover that *or* behaves precisely like *nand* if one permutes two lines on the truth table.

The T^* and F^* translation urns are connecting new information sources to old dispositions. Just as in the two-sender self-assembling signaling model that we started with, the mechanism that structures the new game here is one that forges new saliences by identifying successful sources of information. The translation urns make the truth values of the new propositions P^* and Q^* salient to the old sender dispositions.[18]

5. Modular Composition

Just as agents may self-assemble a simple game by reinforcement on successful sources of information, the same mechanism allows agents to self-assemble more complex games from simpler games. And they may do so more efficiently than they might have evolved the complex game from scratch. We will consider the evolution of a complex logical operation by modular composition as an example of such evolving saliences.

---

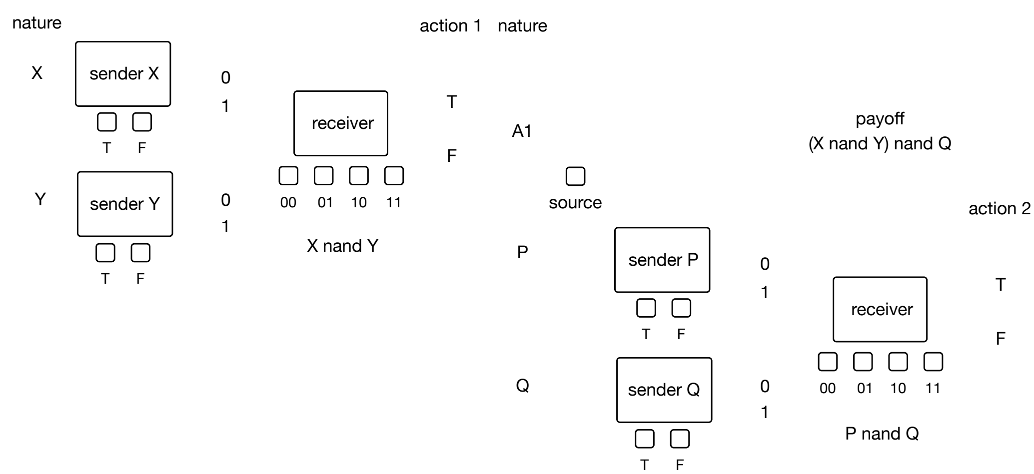
[17] The same mechanism might appropriate the old *nand* rule to a new *nand* context just as fast. The point of considering how an old *nand* rule might be appropriated to a new *or* context is just to illustrate the flexibility of the appropriation dynamics.

[18] Rather than simply fixing the new propositions P^* and Q^*, one could also give the translation urns a source urn to identify potential sources of information to translate. That would be a rather more subtle reinforcement network.

On this model, we start with two logical modules that have each evolved to compute *nand*. The first computes X *nand* Y from the truth values of X and Y, and the second computes Z *nand* Q, where the truth value of Q is from nature and Z initially has an equal chance of being the output of the operation X *nand* Y (labeled $A1$ in the figure) and the randomly-determined, unbiased truth value of proposition P from nature. On each play of the game, sender P draws a ball from her source urn to determine whether to condition her action on $A1$ or on the truth value of proposition P. The initially even probabilistic disposition might be represented by there initially being one ball of each source type $A1$ and P in sender P's source urn.

The composite system is successful on a play if and only if it computes (X *nand* Y) *nand* Q. If the composite system is successful, sender P reinforces on the successful source by adding a ball to her source urn of the type drawn on the play. This is the only urn that gets updated. To focus on the evolution of salience, we again suppose that the modules are relatively stable and just consider how sender P's information source evolves under simple reinforcement learning.

For the composite system to be routinely successful, the P sender must evolve successful saliences. Specifically, she must learn to use $A1$, not P, as her information source. Learning to do so will connect the dispositions of the two logical modules to yield a composite system that computes (X *nand* Y) *nand* Q. Reinforcement on sources that in fact lead to successful action might be thought of as the glue of modular composition.



On simulation, sender P quickly learns to use $A1$, not P, as her source. With 1000 runs, a cumulative success rate better than 0.95 occurs 0.88 of the time on 10^6 plays, 0.69 of the time on 10^4 plays, and 0.51 of the time on 10^3 plays. In contrast, if one tries to evolve the complex operator from scratch by simple reinforcement learning, a cumulative success rate better than 0.95 occurs only 0.23 of the time on 10^6 plays, 0.02 of the time on 10^4 plays, and 0.00 of the time on 10^3 plays.

Consider the same payoffs, but suppose that both sender P and sender Q begin the game with uncertain sources. Specifically, sender P begins with a source urn representing an equal probably of using P or $A1$ and sender Q begins with a source urn representing an equal probability of using Q or $A1$. And, again, we suppose that the dispositions of the two senders are initially flexible and model the even probabilistic dispositions for each source with one ball of each source type in each source urn. And we again suppose that they learn which sources are successful by simple reinforcement. In order to be successful on the task at hand, the senders must coevolve so that sender P typically uses $A1$, not P, and sender Q typically uses Q, not $A1$.

On simulation, sender P nearly always learns to use $A1$ as her source and sender Q nearly always learns to use Q as her source. While they typically co-evolve these successful dispositions somewhat slower than when just sender P is uncertain regarding her source, they evolve them much faster than the complex operator might evolve from scratch by simple reinforcement learning. With 1000 runs, a cumulative success rate better than 0.95 occurs 0.72 of the time on 10^6 plays, 0.34 of the time on 10^4 plays, and 0.15 of the time on 10^3 plays.

The greater the number of inputs to a template and the more possible sources, the harder it is for the senders to evolve successful saliences under simple reinforcement learning. That said, the present examples illustrate that evolving new saliences might accomplish the task faster even when there is significant uncertainty regarding what successful saliences would look like.

This is particularly notable here inasmuch as we are using only simple reinforcement learning, a very slow, low-rationality dynamics. Other learning dynamics allow agents to evolve successful saliences much faster and more reliably than simple reinforcement.[19] And some of these are good models for human learning in specific contexts.[20] But such learning dynamics also allow successful dispositions to evolve from scratch much faster than by simple reinforcement learning. To show the virtues of template transfer and modular composition, one must compare the speed of these evolutionary processes against evolution from scratch using the *same* learning dynamics for both.

Given a particular learning dynamics, the situation here mirrors human problem solving. Sometimes one happens upon a successful analogy between the problem at hand and a similar problem that one has already solved as in the template transfer game above. Other times one may happen upon a way to compose solutions to simpler problems into a solution to a more complex problem by recognizing relevant saliences as in the present case of modular composition. In such cases, one may solve the problem at hand more efficiently than by solving it from scratch. But in the absence of such serendipity one must rely on slower, sometimes less sure, evolutionary processes.

[19]Such learning dynamics include ARP learning, bounded reinforcement with punishment, and win/stay-lose-randomize with reinforcement. These dynamics work well even when there are many potential sources of information to check. See (Barrett and Zollman [2009]; Barrett *et al.* [2017]) for discussions.

[20]ARP learning, for example, captures the observed behavior of animal and human subjects in some contexts, and bounded reinforcement learning with punishment has a natural analog in prediction-expectation learning governed by dopamine neurons. See (Schultz *et al.* [1997]) for the biology, (Bereby-Meyer and Erev [1998]) for the original description of ARP learning, (Barrett and Zollman [2009]) for an extended discussion of ARP learning, and (Cochran and Barrett [2020]) for data regarding how human agents in fact learn in the context of signaling game.

6. Discussion

The models here illustrate how generalized signaling games might self-assemble even as the agents coevolve optimal dispositions for playing the coevolving game. The mechanism for the evolution of the game is network formation by simple reinforcement on information sources that in fact lead to successful action. The network forms as the agents' saliences coevolve with their dispositions to process the information to which they attend. When successful, such games might be understood as self-assembling rule-following abilities or concrete implementations of algorithms, and such capacities might be understood as performing basic epistemic functions like representation, communication, and inference.

We also considered how a pre-evolved template might be appropriated to a new context by reinforcement on successful information sources and how old evolved sender dispositions might come to be triggered by new sources of information. This latter case is another example of self-assembly by means of simple reinforcement on successful sources of information, but it works in the other direction. When successful, the translation agents learn which old dispositions might be salient to new sources of information. And, when they do, they evolve an analogy between the old stimuli that evolved the dispositions and the new sources of information.[21]

In the models we have considered the initial contents of the source urns have been fixed. This means that we are hardwiring the *potential* sources of successful information the agents will check. A natural extension would be to allow agents to investigate novel sources information under a learning dynamics like reinforcement with invention. Here one might imagine a source urn that starts with a single black ball. When the black ball is selected, the agent randomly tries a new source of information. If that play is successful, she adds a new ball type to the urn for that information source. In this way, an agent might find new sources of information when those currently being used are not optimally successful and thus learn for herself by invention which sources of information should be considered salient. This sort of learning dynamics has been well studied.[22]

While we have focussed here on very simple models, the notion of a basic reinforcement network is quite general. Each node in the network has a way to learn (1) what to look at and (2) what to do when it sees something. It is from these basic building blocks that generalized signaling games may evolve and more complicated games evolve from simpler games. Here we have seen that even simple reinforcement learning allows for the evolution of nontrivial games and philosophically compelling capacities.

Acknowledgements

I would like to thank Brian Skyrms for many productive conversations on the topic of this paper and Travis Lacroix, the editors, and two anonymous reviewers for comments on an earlier draft.

[21]Here the translation agents assume an old set of evolved saliences then learn how to tie these to specified new sources of information that may lead to successful action on translation. If successful, they translate the salient information from the new sources to trigger the old evolved dispositions.

[22]See (Alexander *et al.* [2012]) for a discussion of this learning dynamics and (Barrett [2014]; Barrett and Skyrms [2017]) for examples of its effectiveness in application.

Jeffrey A. Barrett
Logic and Philosophy of Science
UC Irvine
Irvine, California, USA
j.barrett@uci.edu

References

Alexander, J. M. , B. Skyrms, S. L. Zabell [2012]: 'Inventing New Signals', *Dynamic Games and Applications*, **2**, pp. 129–45.

Argiento, Raffaele, Robin Pemantle, Brian Skyrms and Stas Volkov [2009]: 'Learning to Signal: Analysis of a Micro-Level Reinforcement Model', *Stochastic Processes and their Applications*, **119(2)**, pp. 373–90.

Barrett, Jeffrey A., Brian Skyrms, and Aydin Mohseni [2019]: 'Self-Assembling Networks', *The British Journal for the Philosophy of Science*, **70(1)**, pp. 301–25.

Barrett, Jeffrey A., Brian Skyrms, and Calvin T. Cochran [2019]: 'On the Evolution of Compositional Language', forthcoming in *Philosophy of Science*.

Barrett, Jeffrey A., Brian Skyrms, and Calvin T. Cochran [2018]: 'Hierarchical Models for the Evolution of Compositional Language', *Institute for Mathematical Behavioral Sciences Technical Report* MBS 18-03. https://www.imbs.uci.edu/research/. Accessed 26 March 2020.

Barrett, Jeffrey A. and Kevin Zollman [2009]: 'The Role of Forgetting in the Evolution and Learning of Language', *Journal of Experimental and Theoretical Artificial Intelligence*, **21(4)**, pp. 293–309.

Barrett, Jeffrey A. and Brian Skyrms [2017]: 'Self-Assembling Games', *The British Journal for the Philosophy of Science*, **68(2)**, pp. 329–53.

Barrett, Jeffrey A., Calvin T. Cochran, Simon Huttegger, N. Fujiwara [2017]: 'Hybrid Learning in Signaling Games', *Journal of Experimental & Theoretical Artificial Intelligence*, **29(5)**, pp. 1–9.

Barrett, Jeffrey A. [2016]: 'Truth and Probability in Evolutionary Games', *Journal of Experimental & Theoretical Artificial Intelligence*, **29(1)**, 219–25.

Barrett, Jeffrey A. [2014]: 'Rule-Following and the Evolution of Basic Concepts', *Philosophy of Science*, **81(5)**, pp. 829–39.

Barrett, Jeffrey A. [2007]: 'Dynamic Partitioning and the Conventionality of Kinds', *Philosophy of Science*, **74**, pp. 527–46.

Barrett, Jeffrey A. [2006]: 'Numerical Simulations of the Lewis Signaling Game: Learning Strategies, Pooling Equilibria, and the Evolution of Grammar', *UC Irvine Institute for Mathematical Behavioral Sciences Technical Report* MBS06-09. http://www.imbs.uci.edu/tr/abs/2006/ mbs06_09.pdf Accessed 26 March 2020.

Bereby-Meyer, Y. and I. Erev [1998]: 'On Learning to Become a Successful Loser: A Comparison of Alternative Abstractions of Learning Processes in the Loss Domain', *Journal of Mathematical Psychology*, **42**, pp. 266–86.

Cochran, Calvin T. and Jeffrey A. Barrett [2020]: 'How Signaling Conventions are Established', manuscript.

Herrnstein, R. J. [1961]: 'Relative and Absolute Strength of Responses as a Function of Frequency of Reinforcement', *Journal of the Experimental Analysis of Behaviour*, **4**, pp. 267–72.

Huttegger, Simon, Brian Skyrms, Pierre Tarrès, and Elliott Wagner [2014]: 'Some Dynamics of Signaling Games', *PNAS*, **111** (Supplement 3), pp. 10873–80.

LaCroix, Travis. [2020]: 'Using Logic to Evolve More Logic: Composing Logical Operators via Self-Assembly', *British Journal for the Philosophy of Science*, forthcoming. doi.org/10.1093/bjps/axz049

Lewis, David [1969]: *Convention.* Cambridge, MA: Harvard University Press.

Pemantle, Robin and Brian Skyrms [2004]: 'Network Formation by Reinforcement Learning: the Long and the Medium Run', *Mathematical Social Sciences*, **48**, pp. 315–27.

Purves, Gordon Michael [2018]: 'Conventionality and Causality in Lewis-Type Evolutionary Prediction Games', *The British Journal for the Philosophy of Science*, forthcoming, https://doi.org/10.1093/bjps/axy076

Schelling, Thomas C. [1960]: *The Strategy of Conflict.* Cambridge: Harvard University Press.

Schultz, W., Dayan, P., and Montague, P. R. [1997]: 'A Neural Substrate of Prediction and Reward', *Science*, **275**, pp. 1593–9.

Skyrms, Brian [2010]: *Signals: Evolution, Learning, & Information*, New York: Oxford University Press.

Skyrms, Brian [2008]: 'Evolution of Signaling Systems with Multiple Senders and Receivers', *Philosophical Transactions of the Royal Society B*, **364**, pp. 771–9.

Skyrms, Brian [2006]: 'Signals', *Philosophy of Science*, **75(5)**, pp. 489–500.